Introduction to Prescriptive Analytics

Introduction to Data Analytics and Value Creation

As discussed in the previous modules, there are four types of data mining (data analytics) as follows:

* **Descriptive:** What happened? Helps uncover valuable insight into the data being analyzed
* **Diagnostic:** Why did it happen? Helps understand the relationships and patterns in the data
* **Predictive:** What is likely to happen in the future? Helps forecast the future behavior of people and markets
* **Prescriptive:** What should I do about it? How should we respond to potential future events based on the analysis? Uses optimization and simulation algorithms to provide guidance and understanding on decisions and answers

This module introduces *prescriptive analytics*, which picks up where predictive analytics leaves off. While predictive analytics gives you a look at the future from the perspective of a fixed point in time, prescriptive analytics continues looking at the future as time moves forward. It does this through the use of operations research and artificial intelligence techniques such as optimization, simulation, and machine learning, which gives it the ability to absorb new data and improve its algorithms over time (potentially without human intervention). Most importantly, prescriptive analytics can recommend future courses of action for business leaders to follow.

# Techniques of Prescriptive Analytics

While the use of prescriptive analytics can be as simple as taking results from *descriptive* and *predictive* analytics, and making business decisions straight away, it can be as complex as developing a sophisticated machine learning algorithm that would tax even the most accomplished data scientist. However, for the purpose of introducing the topic, we’ll focus on two commonly used techniques: *optimization* and *simulation*.

## Optimization

As the name implies, optimization techniques seek an optimal solution to a business problem, which in turn leads to optimal decision making. In general, problems can be expressed mathematically and are subject to a set of constraints, also expressed mathematically, that must be satisfied. For instance, the population of a particular rodent can be described by a mathematical equation that takes into account reproduction and predation (the rate at which rodents are eaten by predators), as well as food availability. When food starts to run out, the population cannot continue growing, so food availability is a constraint that would need to be expressed mathematically and incorporated into the model. Another constraint would be the population of predators, which together with limited food supplies places limits on the rodent population.

One important optimization technique is *linear programming*, in which the mathematical expression for the model is linear and the optimization goal is to either minimize or maximize that expression. This equation is called the objective function and is subject to various constraints including upper bounds, lower bounds, and inequalities. Linear programming is especially useful for blending problems in process industries, production planning in manufacturing, cash flow matching in finance, and other applications.

Another optimization technique is *integer programming,* which is similar to linear programming in that it also seeks to minimize or maximize an objective function subject to various constraints. However, in this case the variables are limited to integer values. Integer programming is useful for portfolio optimization in finance, dispatching electric generating units in energy production (called unit commitment), scheduling and routing in transportation, and in various supply chain applications among others. A noteworthy application of linear programming in business is the setting of production and inventory levels to meet forecast demand at given sales locations.

### Identifying the Goal

All optimization problems have several common elements:

* **Decision variables**: The variables whose values the decision-maker can choose; the goal of linear programming is to find the specific values of the decision variables that optimize the result.
* **Objective function:** The mathematical equation to be optimized in terms of finding a maximum or minimum value (e.g., lowest cost, lowest weight, maximum profit, etc.).
* **Constraints:** Physical, logical, or economic realities (expressed mathematically) that limit the values the decision variables can take.

As a simple example, consider a factory that has the objective of maximizing the net income it generates through the manufacturing and shipping of products. The objective function would then mathematically represent the gross revenue minus the cost of operating the factory, which includes a monthly fee incurred for renting warehouse space to store excess product before shipping. (We’ll keep things very simple by ignoring other expenses.) The constraints are thus the largest number of products that can be manufactured in a given amount of time (say, a day) and the rate at which products can be shipped during that day. If the factory’s production rate exceeds the rate at which products can be shipped, the excess product will accumulate and need to be stored in the rented space.

The factory’s options are as follows:

* Reduce production so that all product units can be shipped without needing to rent warehouse space. The problem with this is that it reduces the potential income that the factory could be earning. (We’ll assume that demand for the product is very high.)
* Ensure that production is always maximized even if warehouse space is needed. After all, more products mean more sales, which means more income. The problem with *this* is that high warehouse rent might cut into the potential income the factory could generate.
* Find an optimal, “in-between” point using linear programming. Your goal is to find the production rate that is high enough to generate good revenue but not so high that the warehouse rental costs begin driving net income down. Somewhere in between the two extremes lies the optimal solution that maximizes net revenue.

Needless to say, linear programming can indeed be an extremely useful tool for these complicated kinds of problems. You can imagine the complexity of a “real” problem with tens or hundreds of decision variables and an equal number of constraints. Thanks to computers, these problems are entirely solvable.

In the previous example, the goal was to maximize net income. One can also set up a linear programming problem to *minimize* an objective, such as minimizing cost, weight, or waste.

For the mathematically interested reader, here is the form in which a typical linear programming equation and its constraints are expressed:

Objective function: max *( x1+ 6x2 )*

Subject to (constraints): 0 ≤ *x1* ≤ 200

0 ≤ *x2* ≤ 300

*x1* + *x2* ≤ 400

In this case, *x1+ 6x2* represents a quantity to be maximized and the three constraints specify the range of values that the two decision variables can take.

In a real business problem, the mathematical solution of a linear programming equation is seldom the final step. It’s usually advisable to perform a sensitivity analysis to examine how the optimal solution might vary if any of the inputs are changed. In some cases, input values may have random components, in which case the problem is solved by *stochastic* optimization.

In conducting a “what-if” exploration, worst-case scenarios are produced by using the most pessimistic input values (e.g., an inefficient production line and exorbitantly high warehouse rents for our earlier factory example) and best-case scenarios are produced by using the most optimistic input values (e.g., a state-of-the-art, high volume production line and unlimited free storage for excess production).

## Simulation

The second prescriptive analytics technique we’ll discuss (albeit more briefly than our discussion of linear programming) is *simulation*. A simulation model imitates real-world situations mathematically and takes uncertainty into consideration. In other words, the input variables aren’t necessarily fixed but may have ranges or distributions that reflect real-world uncertainty. Thus, simulation produces a range of output values rather than a single, bottom-line answer. This approach is amenable to scenario planning, in which a company uses simulation models to produce a range of possible outputs that correspond to different operating scenarios. Simulation models are also useful for sensitivity analyses, which allow decision-makers to see how sensitive a system is to changes in operating conditions. Perhaps best of all, simulations conducted with sophisticated computers make it easy to answer “what-if” questions on the developed computer model without having to build real (physical) systems on which to experiment.

## Experimental Design

Regardless of the simulation approach employed, meaningful and actionable results depend on an appropriate *experimental design*. The length of a simulation, the selection of decision variables, and the interpretation of modeling results must all be carefully considered before you can begin analyzing a model’s output. For instance, one is normally interested in the steady-state results for a system—how the output “settles down” after random short-term fluctuations die out. This requires knowing when to discard output results (e.g., during the initial stages) and when the model has reached steady-state conditions.

# Steps for Applying Prescriptive Analytics

The process of applying prescriptive analytics to a business problem has several major components. First, you want to identify the project goals and follow the steps below.

* Determine the questions to be answered
* Identify scenarios to be investigated
* Identify decision objectives, criteria, and parameters
* Determine the needs of the end-user
* Determine data requirements
* Determine the hardware, software, and computational requirements
* Prepare a time plan
* Develop a cost plan and billing procedure

Next, you want to test the models or (at least) describe the process for testing the models to ensure that they provide meaningful recommendations. This is imperative for assessing and preventing costly mistakes.

# Business Challenges

A significant business challenge when using prescriptive analytics is truly understanding and acting on your data to optimize the decisions based on the prediction or optimization models. For example, to increase profit margins, you may need to increase sales or reduce prices, optimize inventory levels, reduce expenses, or explore other potential solutions and models to identify opportunities for improvement.

## Understanding Your Findings

After creating the predictive/prescriptive model from your data, you need to interpret your results. You must use optimization and simulation algorithms to guide the answers you find and decisions you make. You then must express or communicate your data model in several ways, including using only words, creating a table or chart, expressing score values, or establishing key performance metrics. You are then able to use the results of your data analysis to decide upon the best sequence of action as suggested by your predictive and prescriptive model. You know your business; you are familiar with the business problem and historical data, which means that you are the subject matter expert who can accept or reject the model and its suggestions.

## Strategy Decisions

Developing prescriptive analytics requires answering several critical strategic questions, including the following:

* Is a prescriptive model needed?
* If a model is needed, which one?
* If a model is in place, should it be revised?
* If a model is in place, should it be retired?

Make strategic decisions and plans, and always revisit the model assumptions and validate them with new current data.